The Global Distribution of Biological Nitrogen Fixation in Terrestrial Natural Ecosystems

Davies-Barnard, T.[1], Friedlingstein, P.[1]

[1] College of Engineering, Mathematics and Physical Sciences, University of Exeter, Laver Building, North Park Road, Exeter, EX4 4QE, UK

Corresponding author: T Davies-Barnard (t.davies-barnard@exeter.ac.uk)

Key Points:

- Evapotranspiration and productivity (e.g. NPP) are unreliable predictors of terrestrial biological nitrogen fixation at the global scale.
- Free-living biological nitrogen fixation makes up at least a third of the terrestrial total.
- Global terrestrial biological nitrogen fixation is likely in the range of 52 – 130 TgN $year^{-1}$.

Abstract

Biological nitrogen fixation is a key contributor to sustaining the terrestrial carbon cycle, providing nitrogen input that plants require. However, the amount and global distribution of this fixation is highly disputed. Using a comprehensive meta-analysis of field measurements we make a new assessment of global biological nitrogen fixation (BNF). We assessed the relationship between BNF in natural terrestrial environments and empirical predictors of BNF commonly used in terrestrial ecosystem and earth system models. We found no evidence for any statistically significant relationship between BNF and evapotranspiration and net or gross primary terrestrial productivity (NPP or GPP). We assessed the relationship between BNF and 11 other climate variables and soil properties at a global scale. We found that all the variables we considered had little predictive power for BNF. Using averaged biome values upscaled we calculated the median global inputs of BNF in natural ecosystems as 88 TgN yr^{-1}. The range (52 – 130 TgN yr^{-1}) encompasses most recent estimates and broadly agrees with recent independent top-down estimates of BNF. The global values indicate a significant role for free-living, as opposed to symbiotic, BNF, accounting for at least a third of all BNF. This work provides a new global benchmark and spatial distribution dataset of BNF using a bottom-up methodology.

1 Introduction

The terrestrial carbon cycle is an important contributor to the uptake of atmospheric carbon, removing about a third of anthropogenic carbon emissions from the atmosphere (Le Quéré et al., 2018). Carbon fixation in the terrestrial biosphere is dependent on chlorophyll, of which nitrogen is a key component. But while supply of nitrogen (N) is critical, inorganic N is water soluble and therefore is prone to being washed out of soils (Davis, 2014) or lost via gaseous pathways (Lenhart et al., 2015). One of the key questions for future projections of terrestrial carbon uptake is to what extent N will be available to enable increased growth under high carbon dioxide conditions (Davies-Barnard et al., 2015; Zaehle, et al., 2014). Biological nitrogen fixation (BNF) is one source of new N represented in models. These models are dependent on knowing what the current supply of BNF is, how it is spatially distributed, and what are the environmental drivers of BNF. Models not only need process-based understanding, but global datasets to test the model on. This work aims to provide the latter.

Two primary classes of BNF can be distinguished: symbiotic (also known as associative, or nodulating fixation) (Granhall, 1981), and free-living (also known as non-symbiotic or asymbiotic) (Reed et al., 2011). Both are found in ecosystems worldwide, to a lesser or greater extent. Symbiotic BNF can be defined as fixation in association with higher plants in the form of root nodules (Granhall, 1981). Fixing plants are mostly part of the legume family (Fabeaceae) and include a full span of sizes and forms, from small plants like clover (Trifolium), shrubs such gorse (Ulex), to trees such as alder (Alnus). Questions exist about to what extent symbiotic N_2 fixing plants are facultative or obligate fixers, and estimates of fixation can vary accordingly (Menge et al., 2009; Sheffer et al., 2015).

Free-living fixation encompasses a huge range of organisms in virtually all parts of terrestrial ecosystems. In vegetated ecosystems free-living fixation can be found in soil, litter, woody debris, and plant canopies, as well as in bryophytes (mosses) and lichen (Reed et al., 2011). Even where vegetation is sparse, fixation is found in cyptogamic covers (Elbert et al., 2012). Valid discussions surround the usefulness of free-living and symbiotic as categorisations, as they are neither consistent nor discrete. BNF associated with bryophytes has been shown to be symbiotic (Adams & Duggan, 2008) but is usually classified as free-living, and free-living fixers are more phylogenetically diverse than makes a logical grouping (Reed et al., 2011). Use of these well-known classifications is a helpful shorthand for us in this context, but there is much more nuance to this issue than we present here. The relative contribution of symbiotic and free-living fixers to global BNF is an ongoing debate that we aim to shed some light on.

Existing global terrestrial estimates of BNF from natural sources provide a substantial range with little sign of consensus emerging over time. There are a range of different methods, which can be roughly categorized into three groups. Some global BNF estimates are top-down estimates using a N budget (e.g. Vitousek et al., (2013)), which takes known global values of carbon to nitrogen ratios and ^{15}N and uses these to calculate the required BNF. Other top-down methods can be less empirical global budgets (e.g Delwiche, (1970)). There are number of ‘bottom-up’ estimates, of which Cleveland et al., (1999) is the best known. These use a meta-analysis of available field measurements then upscale biome averages to a global total. The majority of global BNF values are models or model and data combinations. These include field data in conjunction with models (e.g. Wang & Houlton, (2009)), new data-informed models (e.g. (Xu-Ri & Prentice, 2017)), and existing models with new predictive data (e.g Galloway et al., (2004)).

The seminal meta-analysis of BNF done by Cleveland et al., (1999) established an empirical relationship between evapotranspiration (ET) or net primary productivity (NPP) and non-agricultural BNF. This relationship has been used by many terrestrial carbon-nitrogen models (Bloh et al., 2018; Goll et al., 2017; Koven et al., 2013; Smith et al., 2014) as well as other estimates of total global BNF (Cleveland et al., 2013; Galloway et al., 2004). A fuller description of the range of common model calculations of BNF can be found in Zaehle, et al., (2014) or Meyerholt et al., (2016). As the most data-based BNF estimate available, the global spatial distribution of BNF based on Cleveland’s model of ET is sometimes used as ‘observations’ for comparison with model output (e.g. Meyerholt et al., (2016)). Therefore, the reliability of the relationship between NPP and ET and BNF is important to accurate modelling. Given that 20 years of new field measurements are now available and there is continued uncertainty about the global total and spatial pattern of BNF, a new ‘bottom-up’ assessment is timely.

This paper aims to give a new comprehensive insight into the empirical relationship between BNF in natural ecosystems and a range of related variables. We consider linear modelling to establish the relationship between BNF and soil and climate variables. We also use a upscaled biome approach using land cover groupings to provide global total and spatial distribution BNF estimates based only on measured data. We conclude by comparing our new global calculations to previous assessments of global BNF.

2 Methods

We reviewed over 300 papers and books and collected information about the N_2 fixation, fixer type, latitude and longitude, and vegetation type. This gave over 550 entries. We exclude some measurements, including some used in previous studies, because they do not meet our standard for reliability. Our overarching principles for inclusion are:

- The measurement must be stated by the author in annual units. Therefore, values that are per hour or per day or representative of short-scale measurements, are not scaled up to annual estimates and used here. We do include scaled-up measurements if the author has themselves calculated an annual value, as we assume the author judges the measurements to be sufficiently representative. Where a measurement is given for the entire growth season, we include it as representative of the whole year.

- Values must be in comparable units of N. Values only given in C_2H_4 (ethylene) or C_2H_2 (acetylene) are therefore excluded because conversion between C_2H_4 and N is variable (Ley & D'Antonio, 1998; Nohrstedt, 1985; Saiz et al., 2019). However, where the author has made the conversion, we accept their scientific judgement.
- The measurements must be representative and not anomalous. Measurements that specify that they are the maximum, represent an uncommon soil or vegetation type, are noted by the authors as being unreliable, or similar provisos, are excluded from the analysis.
- Values must be from the primary source. The practice of using numbers cited in reviews, other secondary material, or from unpublished data increases the risk of transcription errors. Therefore, we only include values verifiable in the primary source. For that reason, we include in our dataset the precise location (e.g. page number) of the data within the source.
- Measurements must distinguish the source of BNF to some extent. Being unable to specify the source of the BNF is suggestive of unreliable methods, for instance budgets that assume an amount of BNF. Where the BNF sources are not differentiated it is possible that issues such as including non-biological nitrogen fixation (e.g. weathering), could be present, overestimating the amount of BNF.
- The values must be field measurements, not 'guesses', 'estimates', or values deduced from carbon or nitrogen budgets. Though we understand the useful role these estimates had in previous work, it is difficult to be sure that they are accurate, particularly as the methods of reaching the estimates are often opaque.
- Values must be weighted by the presence or cover of the plant or organism. Some measurements assume 100% coverage of say, a fixing legume, but do not specify what the level of coverage of that fixing legume is in the environment. This then

requires a highly uncertain assumption of the cover. We could not be sure that any resulting relationship was genuine or due to error in the cover percentage.

This resulted in 253 usable values. A reference list of all sources used in this paper is available in Supporting Information section 1.

These criteria are more stringent than those used by Cleveland et al. (1999), because with increased data availability comes the opportunity to discard less reliable data. The most notable difference between our inclusion criteria and that of Cleveland et al. (1999) is our exclusion of unweighted symbiotic measurements. Cleveland et al. (1999) includes unweighted values by averaging the available coverage percent in that biome. We could have done similarly, however we found only 16 estimates of symbiotic fixing coverage across all biomes, from 12 sources (Baker et al., 1986; Bauters et al., 2016; Blundon & Dale, 1990; Bowman et al., 1996; Cech et al., 2010; Fahey et al., 1985; Grove & Malajczuk, 1992; Johnson & Mayeux, 1990; Kummerow et al., 1978; Menge & Chazdon, 2016; Permar & Fisher, 1983; Rundel et al., 1982), which range from 0.3% (Cech et al., 2010) to 34% (Rundel et al., 1982). Given this range and the small sample size, we could not be certain that wrong assumptions of symbiotic coverage would not skew the results. Therefore we only include values where the ecosystem average is given.

The method of measurement is also a significant issue for BNF. We exclude all ‘budget’ type estimates of BNF, where the BNF is extrapolated from measurements of say, large scale deposition, nitrogen uptake, and nitrogen leaching. In principle, we include all direct measurements, however the method of measurement may have some effect on the resultant values. There are two main methods of measuring BNF: the acetylene-ethene reduction assay method (Hardy et al., 1968) (ARA) and the $^{15}N_2$ method.

ARA works on the basis of the enzyme mainly responsible for fixation having a preference for acetylene (C_2H_2) over N_2. The amount of resultant ethylene (C_2H_4) indicates fixation, and can be converted to the equivalent amount of N_2 fixed. The conversion factor of 3:1 C_2H_4:N_2 is commonly assumed (Hardy et al., 1968), and was used by Cleveland et al., (1999) to convert values not already reported in units of N. But this conversion factor varies considerably, with studies suggesting anything from 1.6:1 to 5.6:1 (Nohrstedt, 1985) and 1:1000 to 5.363:1 (Saiz et al., 2019) with variation over space, time, and species. This disparity is the reason we do not convert C_2H_4 or C_2H_2 measurements to N. Since ARA method measurements makes up a significant proportion of the measurements available and many studies have site-specific conversion factors or use more than one method for verification, we include them.

The ^{15}N method involves measurements based on known ratios of the stable isotope ^{15}N to the more common ^{14}N. Measurements can use enrichment of soil with ^{15}N or naturally occurring differences. Compared to ARA, ^{15}N methods are more expensive, but thought to be more reliable. Other methods, such as N accumulation within plants, generally don't provide estimates that are acceptable given the other restrictions listed above, but have been included where appropriate.

From the stated vegetation type we matched to the most appropriate IGBP (International Geosphere-Biosphere Programme) Land Cover Type Classification, as used in the MODIS land cover product from Friedl et al., (2010), shown in Table 1. We acknowledge that allocation of a vegetation type to an ecosystem is unavoidably a normative judgement.

Table 1. The IGBP Land Cover Type Classifications and corresponding abbreviations used.

Abbreviation	**Name**	**Area (km^2)**	**Number of values. Notes**
ENF	Evergreen Needleleaf Forest	3,849,855	65.
EBF	Evergreen Broadleaf Forest	14,136,082	57.
DNF	Deciduous Needleleaf Forest	1,516,648	3. For BNF types where no values for DNF are available the BNF value of ENF is allocated to DNF.
DBF	Deciduous Broadleaf Forest	1,195,671	22.
MF	Mixed Forest	10,233,122	8.
Shrub Cl	Closed shrublands	47,447	0. No BNF values for Shrub Cl are available so the BNF value of Shrub Op is allocated to Shrub Cl.
Shrub Op	Open shrublands	21,312,930	22.
Sav Wood	Woody savannas	10,187,798	8.
Savanna	Savannas	9,649,685	14.
Grass	Grasslands	18,449,115	18.
Wetland	Permanent wetlands	709,907	34.

-	Croplands	11,804,307	Allocated BNF values of Grass
-	Urban and built up	86,447	Allocated BNF values of Grass
-	Cropland/Natural vegetation mosaic	6,200,218	Allocated BNF values of Grass
Barren	Barren or sparsely vegetated	19,047,032	2.
-	Snow and ice	2,974,617	Excluded from analysis.

Each BNF measurement is categorized to only one of the BNF types described in Table 2. These types cover all the major categories frequently found in the literature. Measurements are allocated to the most granular appropriate category and are not duplicated. For instance, if a field study gives 3 values: for free-living, symbiotic, and the free-living and symbiotic BNF combined, the value for combined free-living and symbiotic would be disregarded and only the separate symbiotic and free-living values would be used. Some types of free-living BNF have been grouped for simplicity, particularly the blue-green algae values which are allocated into soil BNF (see Table 2).

Unless otherwise specified, we use a single representative value of BNF from each physical location and BNF type. Where a range rather than a single value is given, the middle value of the range is used. If a range and a 'best estimate' is given and the best estimate is not the middle value, we use the 'best estimate'.

Table 2. Description and abbreviations for different BNF type categories and the number of measurements in each category.

Abbreviation	Description	Number of values
S	Symbiotic values of BNF.	47
FL-ud	Measurements of BNF from an undifferentiated source or mix of sources of all free-living sources in that environment. In these values there is no distinction between the amount of BNF from different sources of free-living BNF. Where different types are measured separately, they are allocated to one of the below appropriate categories. This category includes soil crusts or cryptogamic covers.	71
FL-sl	Measurements of free-living fixing within soil, including green-blue algae and cyanobacteria (unless otherwise specified as being associated with, for instance, moss).	29
FL-lr	Measurements of free-living fixing within leaf litter.	26
FL-ln	Measurements of free-living fixation associated with lichens. This is assumed to be weighted by the area covered by lichens in the environment measured. Where the values are stated or believed to be unrepresentative of the average lichen cover in the environment, these values are excluded.	33
FL-ms	Measurements of free-living fixation in association with bryophytes. This is assumed to be weighted by the area covered by bryophytes in the environment measured. Where the values are stated or believed to be unrepresentative of the average bryophyte cover in the environment, these values are excluded.	26

FL-cy	Measurements of free-living fixation within the canopy, including epiphytes, leaves, tree trunks, and stems.	14
FL-wd	Measurements of free-living fixing within wood on the ground or other woody debris, excluding leaf litter and stems.	7

To create a dataset where the relationships between BNF and climate and soil variables can be explored, we take the latitude and longitude associated with each value. Where the location is specified in latitude and longitude in the source, this is used, and where it is absent the closest point from the description is used. From this we extrapolate the following for each location of a BNF value:

- Mean annual Gross Primary Productivity (GPP) based on FLUXCOM RS+METEO with CRUNCEPv6 climate, average of 2000 - 2010 (Jung et al., 2017; Tramontana et al., 2016)
- Mean annual Net Primary Productivity (NPP) from the Coupled Model Intercomparison Project Phase 6 (CMIP6) model UKESM1 (United Kingdom Earth System Model 1) historical simulation, r8i1p1f2, 2000 – 2010. (Available from ESGF@CEDA, https://esgf-index1.ceda.ac.uk/search/cmip6-ceda/)
- Mean annual temperature from WATCH, average of 1980 – 1999 (Weedon et al., 2011)
- Total annual precipitation from WATCH, average of 1980 – 1999 (Weedon et al., 2011)
- Mean annual incoming solar radiation from WATCH, average of 1980 – 1999 (Weedon et al., 2011)
- Mean annual humidity from WATCH, average of 1980 – 1999 (Weedon et al., 2011)

- Mean annual pressure from WATCH, average of 1980 – 1999 (Weedon et al., 2011)
- Mean annual ET from LandFlux, average of 1989 – 2005 (Mueller et al., 2013)
- Global phosphorus soil distribution (total including inorganic and organic) from ORNL DAAC, NASA Earth Data. (Yang et al., 2014)
- Soil bulk density from ORNL DAAC, NASA Earth Data Regridded Harmonized World Soil Database v1.2 (CLM resolution) (Saxton et al., 1986; Wieder et al., 2014)
- Soil Organic Content (SOC) of the dominant mapping unit ID from HWSD from ORNL DAAC, NASA Earth Data Regridded Harmonized World Soil Database v1.2 (CLM resolution) (Saxton et al., 1986; Wieder et al., 2014)
- Soil clay fraction by percent weight from HWSD from ORNL DAAC, NASA Earth Data Regridded Harmonized World Soil Database v1.2 (CLM resolution) (Saxton et al., 1986; Wieder et al., 2014)
- Soil sand fraction by percent weight from HWSD from ORNL DAAC, NASA Earth Data Regridded Harmonized World Soil Database v1.2 (CLM resolution) (Saxton et al., 1986; Wieder et al., 2014)
- Soil pH in water for the dominant mapping unit from HWSD from ORNL DAAC, NASA Earth Data Regridded Harmonized World Soil Database v1.2 (CLM resolution) (Saxton et al., 1986; Wieder et al., 2014)

Cleveland et al. (1999) used NPP and ET from the Century ecosystem model. We aim to use satellite or measured data wherever possible. For ET, we use the observations-based Landflux data (Mueller et al., 2013). However, for NPP the situation is hampered by data unavailability. At time of writing, the only satellite derived product of NPP, from MODIS, is unavailable due to errors caused by persistent cloud cover biases. Pragmatically, gross primary productivity (GPP) is a very good proxy for NPP, as NPP is GPP minus plant respiration. In models, NPP is generally approximately half of GPP, so there is reason to believe the two are interchangeable for the purpose of relationship with BNF. There are well established observation-based products for GPP, including from Fluxcom, which is the dataset we use here. However, we appreciate the need for some direct comparison with NPP, the most common variable used for BNF empirical relationships. Therefore, in the initial part of our analysis we also assess NPP from one of the CMIP6 earth system models, UKESM1.

3.0 Results Overall Assessment

The global distribution of BNF measurements (Figure 1) immediately shows the paucity of data available. There is a bias towards north America and north-west Europe, with most of values coming from these areas. Central Eurasia, south Asia, and north Africa are particularly poorly represented.

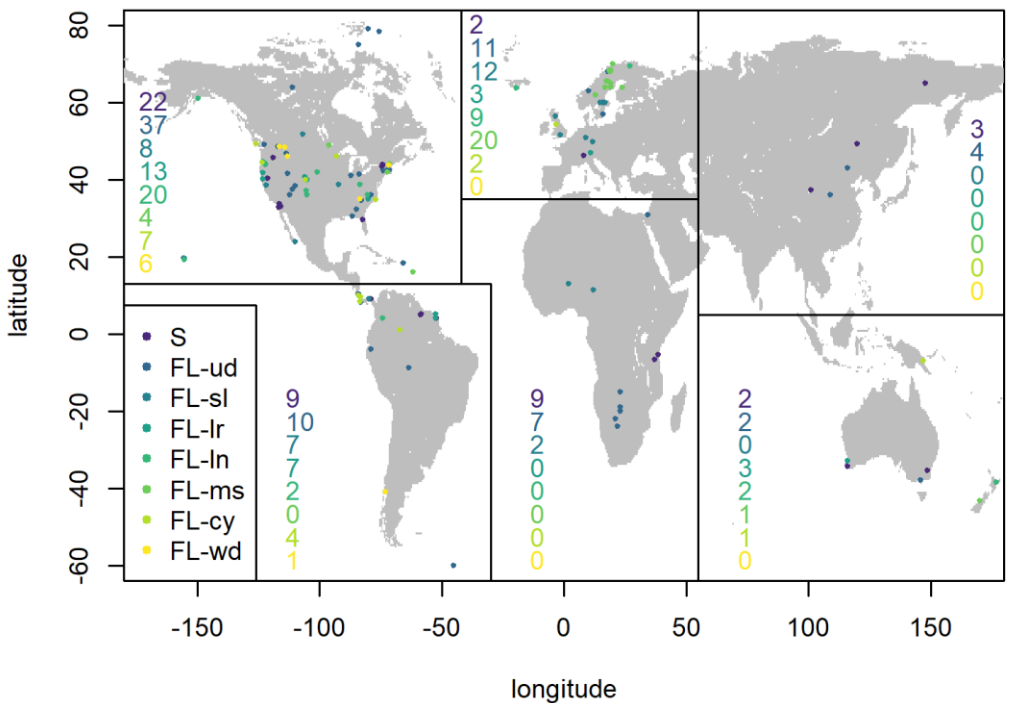
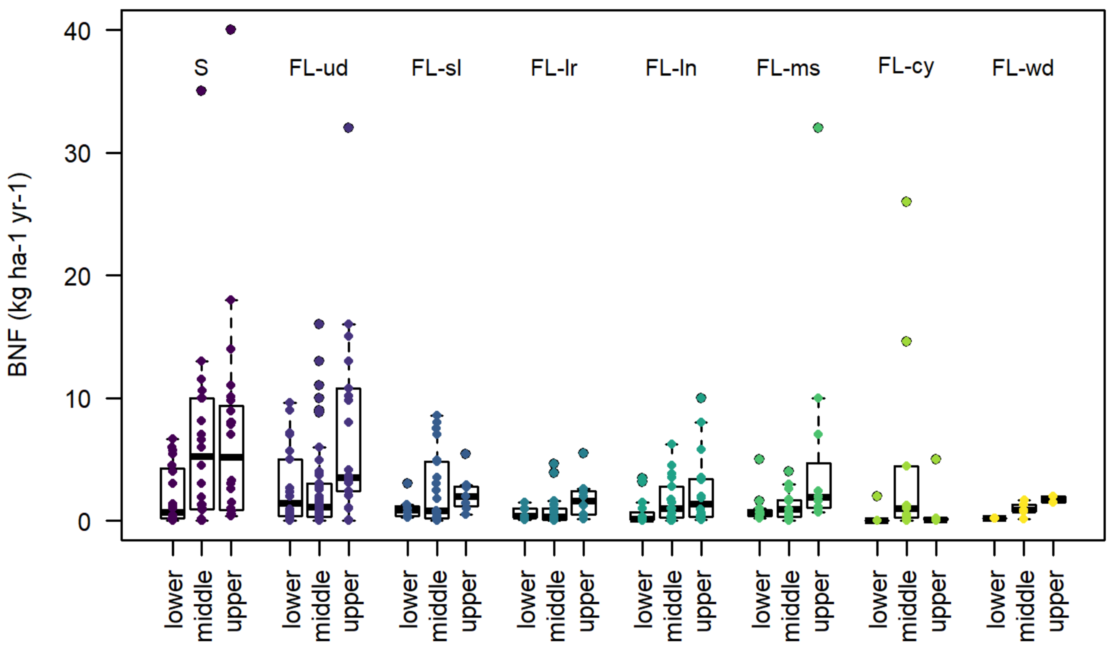


Figure 1. All the data locations mapped, with the colour indicating the type of BNF. Since some regions have multiple measurements associated, less values are shown than are used in other parts of the analysis, as they are over-plotted. The numbers shown in the map relate to the number of values within each category found within the delineated region.

The challenge of this dataset is further revealed by considering the range of reported BNF values. The measurements have been separated here into either values with an upper and lower limit (i.e. a range), or with only one central estimate (Figure 2 a). If the measured values were evenly spread across biomes, one would expect the central single values to fall somewhere between the upper and lower range values. This is not the case in all the BNF types, indicating that the data amount is too low for this pattern to emerge or that the data is very heterogeneous. Only FL-ln (lichens) follows the expected pattern consistently (for range, inter quartile range, and median values). This suggests that any further data acquired could not be entirely relied upon to conform to the pattern of current data, especially in categories with small sample sizes.



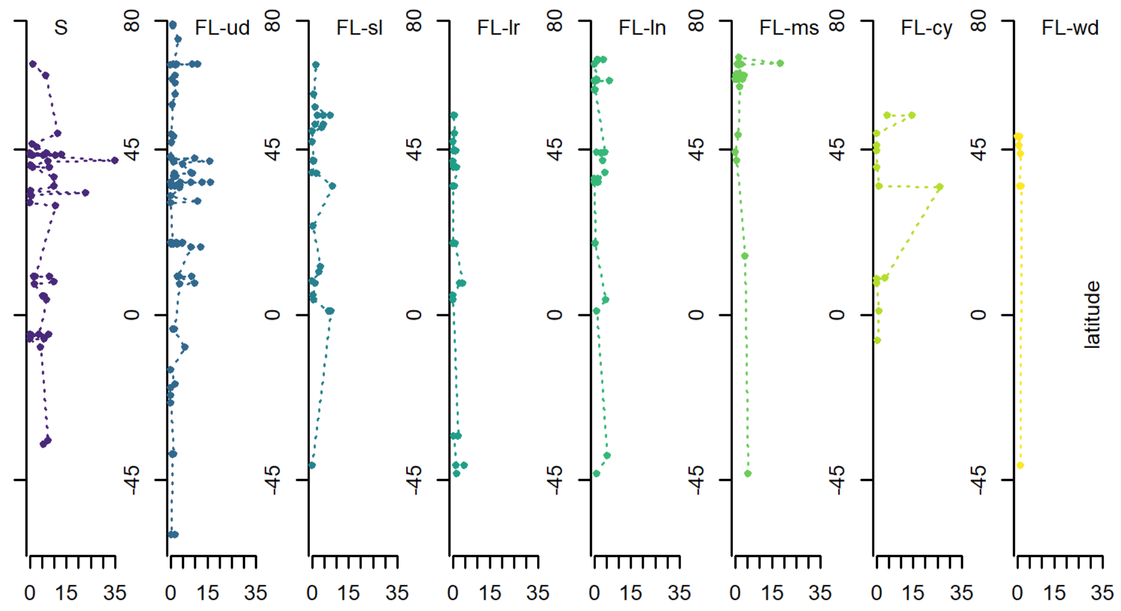


Figure 2. a): Fixation values for each of the BNF types, grouped by lower and upper bounded ranges or a single central value (note a measurement is either represented as a range or as a single value, but not both). For each boxplot, the midline is the median, the upper line third quartile, lower line the first quartile, and the whiskers extend up to 1.5 times the interquartile range from the top of the box to the furthest datum within that distance. Datum beyond 1.5 times the interquartile range are represented as individual points. Overlaid on the boxplots are all the individual points as a ‘beeswarm’ scatter. b): BNF values by latitude and separated by type. These are the central values as described in the methods, i.e. the mean of a range or the single most representative value. BNF units for all cases are kg ha^{-1} $year^{-1}$.

We can see that all the individual free-living BNF categories are relatively small (Figure 2a) compared to symbiotic or FL-ud. The FL-wd category has the smallest range overall and the largest ranges of the free-living categories are FL-cy and FL-ms, which are skewed by a small number of outliers. The FL-ud category is not much higher in mean or median than the individual categories of free-living BNF.

Nearly all the categories have a skew to lower values (the median is lower than the mean) and a high tail. But overall, there is perhaps less difference between the categories than might be expected. The paradigm of more fixation by symbiosis (Cleveland et al., 1999) is difficult to justify looking at these values, as FL-ud and S are both the categories that are highest and have the largest range in values.

Looking at the values of BNF by type across latitudes (Figure 2b), there is a lack of latitudinal pattern that we would expect if productivity or ET were a driver of BNF. There are latitudinal clusters of measurements in the mid latitudes and near the equator, but little evidence that BNF increases with decreasing latitude. Free-living BNF, that might be thought to be higher or more prevalent in cooler climates given high carbon uptake by lichens and bryophytes at high latitudes (Porada et al., 2014), also shows little sign of that trend. S and FL-ud appear to have a peak around 40 °N, but this could be sampling error because of the higher number of measurements around this latitude.

3.1 Statistical Modelling

To assess the relationship between ET and productivity versus BNF we use linear modelling which shows how well correlated two datasets are. If there were a relationship between terrestrial productivity or ET and BNF as strong as Cleveland et al. (1999) found (r^2 = 0.63 for ET), it should be evident in a plot of each value we have versus the GPP (or NPP) or the ET for the nearest grid cell (see methods and Figure 3). However, instead of the positive relationship we might expect, neither ET, NPP, nor GPP show any obvious relationship with BNF overall (Figure 3). For each individual BNF type the pattern is contradictory, with some showing positive, flat, or negative relationships. The r^2 for NPP, GPP and ET versus BNF relationships are low (around zero or negative) and the p-values high (>0.5). Given this, it makes sense to widen the analysis to consider other variables.

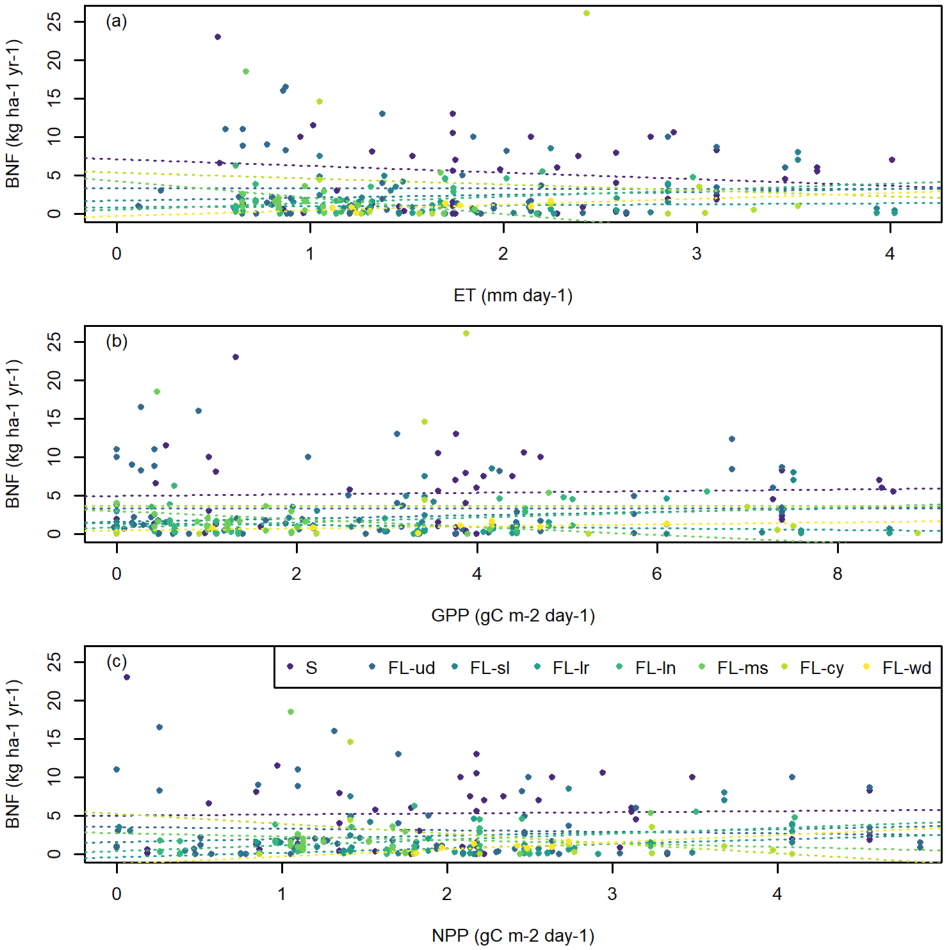


Figure 3. BNF plotted against a) ET (mm day^{-1}), b) GPP (gC m^{-2} day^{-1}) and, c) NPP (gC m^{-2} day^{-1}), colour coded by BNF type (Table 2). Dotted lines represent the linear fit of the BNF type of the same colour.

We expanded our assessment to other variables (listed in the methods) to see which best predict the individual types of BNF (Table 2). We can see the best predicting variable (the highest r^2) of each type of BNF in Figure 4. The category of BNF with the best environmental predictor is FL-wd (r^2 of 0.46, p-value 0.055), although this category has the smallest sample size (7 data, see Table 2). The BNF types with the highest number of measurements have some of the lowest r^2 values (e.g. FL-ud and S). This suggests that we cannot safely assume that if there were a similar number of measurements available for FL-wd the same relationship would be maintained.

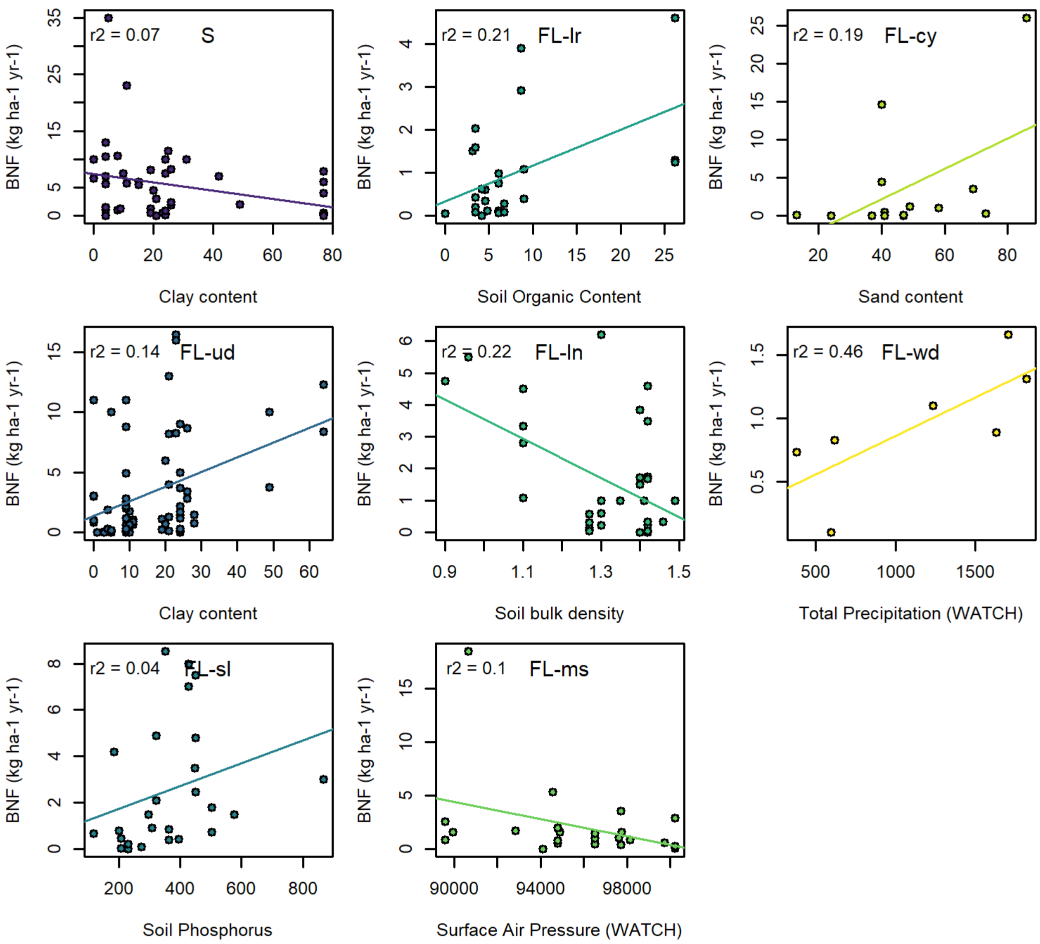


Figure 4 The strongest relationship (highest r^2) between each of the BNF types and the climate or soil variables listed in methods. The linear fit is shown as a line in each of the plots and the corresponding r^2 value is in the top left-hand corner. SOC (kg m^{-2}), ET (mm day^{-1}), Precipitation (mm yr^{-1}), Clay or Sand content (percent), Surface pressure (Pa).

Climate variables in general are not the best predictors of BNF, with only FL-wd (wood) and FL-ms (moss) having the highest r^2 from a directly climate related variable. However, with only 7 data points, FL-wd is not compelling evidence that climate is a key driver. In contrast to the BNF model by Cleveland et al., (1999), none of the types of BNF are best correlated with ET or GPP.

Six of the eight BNF types are best predicted by a soil variable (Figure 4). None of these have a strong predictive power and have correspondingly low p-values. However, in the context of the complete lack of evidence for climatic or productivity controls on BNF, three-quarters of the BNF types being best correlated with soil variables show the most promise for further research.

3.2 Averaged Modelling

Taking the raw values and corresponding soil or climate value does not produce a clear correlation, as seen in section 3.1. Cleveland et al., (1999)'s approach was to bin values using biomes, then build an empirical model from these biome averages. However, there are some limitations with this approach we wanted to avoid, whilst still considering if underlying patterns can be revealed. The subjectivity in the allocation of vegetation types to biomes and the small sample sizes in some biomes are both undesirable. To avoid this, we average across 10 equal sized bins, bounded by the minimum and maximum for the predictive variable range of that BNF type. i.e. the predictive variable or each BNF type the bins are one tenth of the range of the predictive variable. This provides bin averages in a similar way to biome averages, but without the uncertainty and normative judgements about allocations to biome groups. However, because of the low number of values in the free-living categories (excluding FL-ud), we combine these to make a second single free-living value, in a similar approach to Cleveland et al. (1999)’s. We use the sum of all the average of free-living BNF types within each bin to give an alternate FL-ud. We call this FL-sum.

The binning and averaging helps smooth out the variability seen in Figures 1-3 but also reduces the number of values (Figure 5). None of the climate or productivity variables perform well here, though the soil variables are more promising. Sand has the highest r^2 (0.53) for symbiotic (S) BNF but a high p-value (p=0.061). SOC has a lower r^2 (0.38) and higher p-value (p=0.12). The FL-ud category also has a high r^2 for Sand content (0.79) and a correspondingly low p-value (0.0044). Clay also has good predictive power for FL-ud and for FL-sum and S, but the p-values are >0.01 for all.

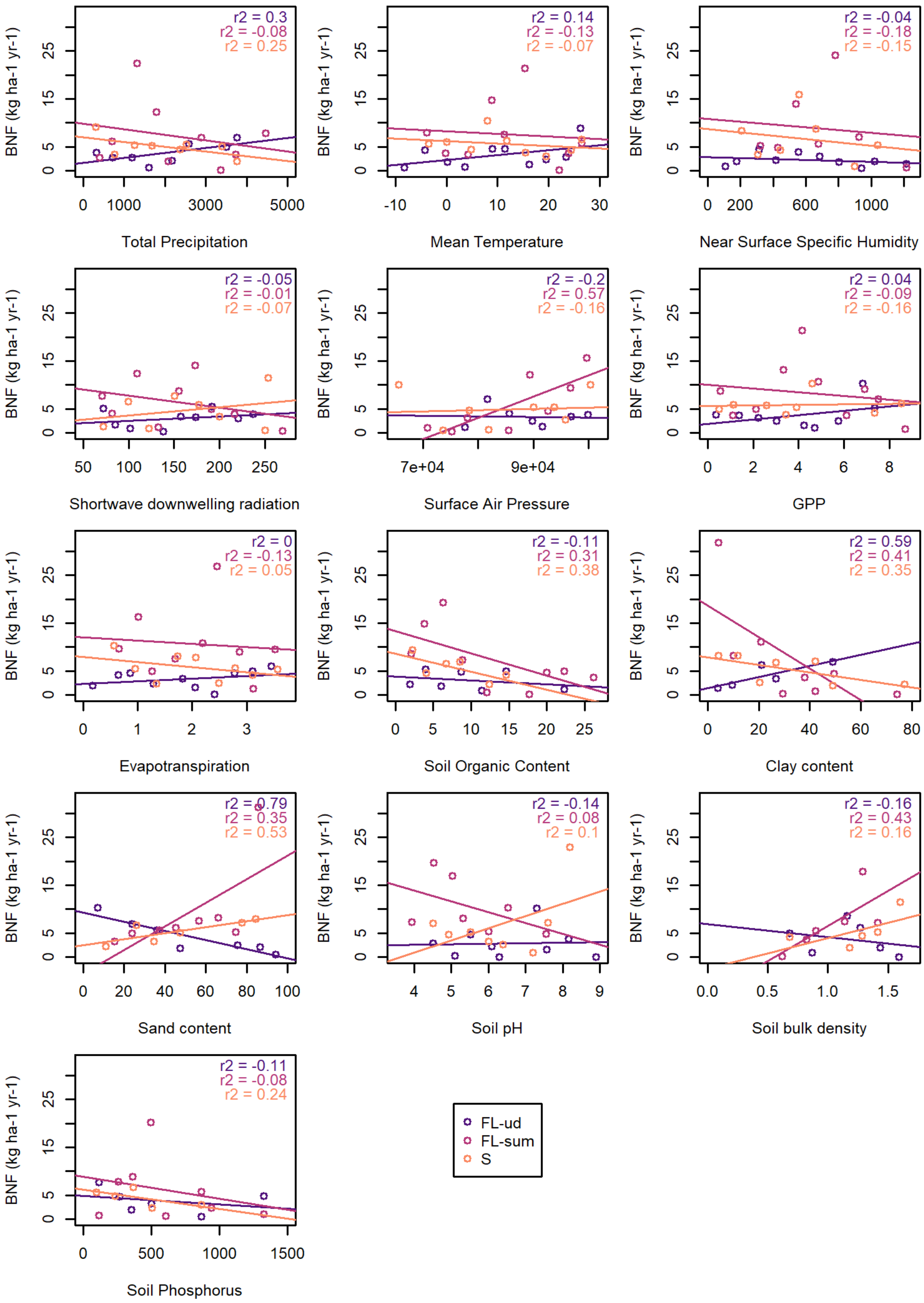


Figure 5. The values and linear model of the binned BNF for S, FL-ud, and FL-sum (see above) is plotted against a range of predictive variables. SOC (kg m^{-2}), GPP (gC m^{-1} day^{-1}), ET (mm day^{-1}), specific humidity (kg kg^{-1}), precipitation (mm yr^{-1}), surface pressure (Pa), surface downwelling shortwave (W m^{-2}), mean annual surface (2m) air temperature (Celsius), total soil phosphorus (gP m^{-2}), soil bulk density (kg dm^{-3}), clay or sand content (percent).

Note that not all-of the 10 bins have BNF values, thus for most variables there are less than 10 data points. The numbers on each plot represent the r^2 for the corresponding colour.

For an empirical relationship to be sound, it seems rational to expect that FL-sum and FL-ud would show the same sign of relationship. Soil Phosphorus, SOC, and Near Surface Specific Humidity are the only variables with consistency of sign and all have a negative relationship (Figure 5). In all these at least one of FL-ud and FL-sum have a negative r^2 value and high p-values. So the relationships are weak, but suggestive that soil properties are more likely to be useful to predicting BNF than productivity or climate.

3.2 Results Mapping

As an alternative to the linear model approach just presented, we also consider a land cover type approach for upscaling BNF similar to that used by Cleveland et al., (1999). We used the allocated IGBP land cover types (see methods section) and upscaled the averaged values (Table 3) to the MODIS map using the same scheme. For this we only consider the FL-ud and S categories. We chose not to attempt (as Cleveland et al. (1999) did) to sum all the different sources of BNF by assuming all sources of BNF in all land cover types for the upscaled measurements in order to increase the robustness of the results. Whereas Cleveland et al., (1999) adds up the average symbiotic and free-living types into a single BNF value, we keep these two separate. This enables us to see which aspects of BNF are contributing to any overall pattern and establish whether the drivers could be different. Excluding FL-sum and using only S and FL-ud also means that standard statistical methods can be used.

Upscaling the biome level mean values shown in Table 3 gives a global total BNF of 102 TgN yr^{-1} (Table 4), with 45% from free-living and 55% from symbiotic. Using the geometric mean that accounts for the lognormally distributed data (Parkin and Robinson, 1993) as used by Cleveland et al. (1999) the global total is 67 TgN yr^{-1}, with 33% free-living BNF. Due to the large range and small number of values available, the mean +/- one standard deviation gives negative values in some cases, particularly the lower free-living estimates. An alternative way of looking at the spread and average of the values is to use the median and interquartile range (see Table 4 and Figure 6 and 7), and this is the approach we focus on.

Table 3 Mean, median, and geometric mean of BNF (kg m^{-2} yr^{-1}) for each IGBP land cover type and the number of measured values used for each.

	ENF	**EBF**	**DBF**	**MF**	**Shrub Op**	**Sav Wood**	**Sav**	**Grass**	**Wetland**	**Barren**
S mean	3.83	5.12	8.30	7.47	0.55	11.63	0.35	6.71	6.92	-
S Geo. mean	1.27	4.32	8.12	7.47	0.55	8.99	0.26	4.85	3.27	-
S median	0.55	5.75	8.30	7.47	0.55	7.90	0.34	8.10	1.50	-
FL-ud mean	1.28	4.25	0.73	0.50	5.89	0.65	0.59	6.06	2.40	0.68
FL-ud Geo. mean	0.69	2.60	0.47	0.45	2.69	0.05	0.20	2.92	1.06	0.68
FL-ud median	1.10	3.11	0.3	0.50	2.77	0.02	0.13	5.00	1.45	0.68

S (number of obs.)	3	13	2	1	1	3	8	5	11	0
FL-ud (number of obs.)	10	16	3	2	16	4	6	5	8	1

Table 4. Global totals of BNF (in TgN yr^{-1}) from symbiotic and free-living sources and in percent of the total.

Global total	**Symbiotic (S)**		**Free-living (FL-ud)**		**S + FL-ud**
Mean	56	56%	45	44%	101
Geometric mean	44	66%	22	34%	67
Median	57	64%	31	36%	88
25% Quartile	31	59%	21	41%	52
75% Quartile	66	51%	66	49%	130

The range of values within categories varies and how much influence this has depends on the extent of that land cover globally. The Grass category is variable for both S and FL-ud but important, as large areas of Eurasia and North America are categorized here as Grass as a proxy for cropland (which we do not directly consider). Open Shrub and Woody Savanna also have disjoints between the mean, geometric mean, and median because of their large range and small sample size. Because of their large coverage Grass and Shrub Open contribute the most to the differences between global totals for the statistical methods. Conversely, Wetland has the largest range, with a substantially higher arithmetic mean than median for symbiotic BNF. But since Wetland covers a small area (see Table 1), this does not significantly affect the global values.

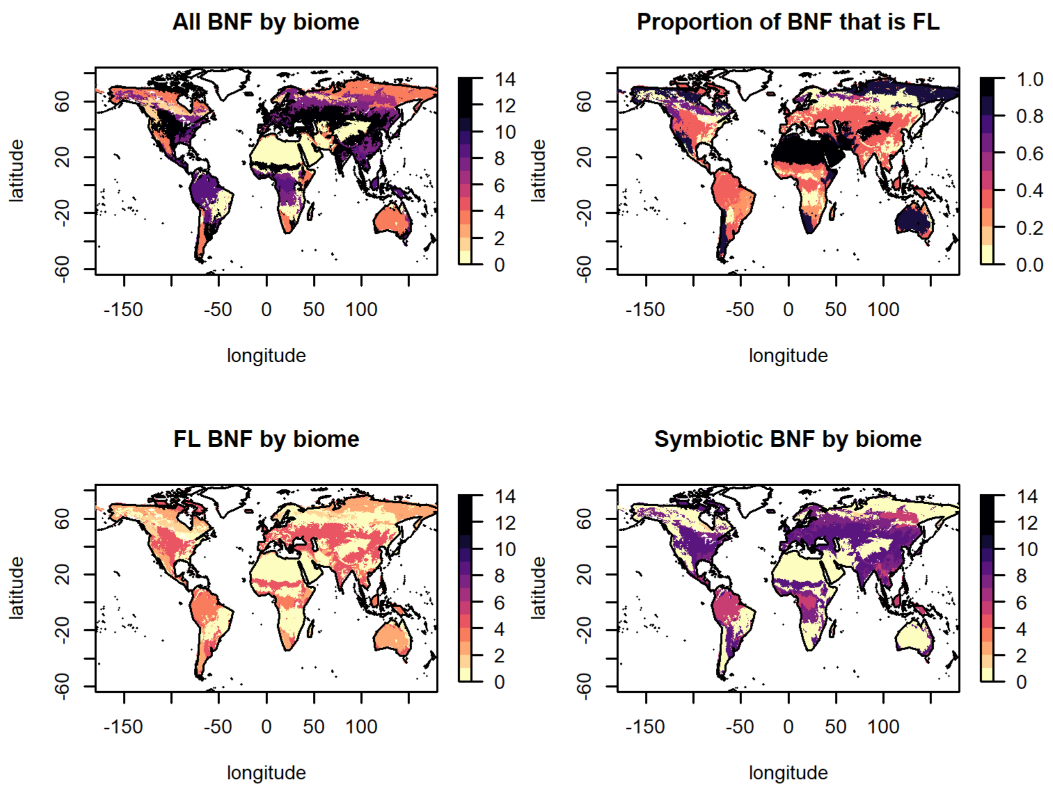


Figure 6. Maps of BNF using median BNF allocated to IGBP land cover types. Top row: Symbiotic (S) and free-living (FL-ud) BNF combined (left) and relative proportion of free-living (FL-ud) BNF (right). Bottom row: Free-living (FL-ud) BNF (left) and Symbiotic (S) BNF (right). BNF in kg ha^{-1} yr^{-1}. The proportion of free-living BNF is 0 (all symbiotic) to 1 (all free-living).

Globally, free-living is consistently smaller than symbiotic BNF, but still a major contributor. The proportion of free-living BNF is between 34 – 48% in any of the statistical values. The balance of symbiotic to free-living BNF also varies regionally (Figure 6) and similar differences lie between symbiotic and free-living BNF as between different biomes. In broad terms, areas absent of symbiotic BNF have higher free-living BNF and vice versa. The exception is EBF and Grass which have relatively high levels of both symbiotic and free-living fixation, resulting in the highest BNF areas. In Figure 6 and Table 3 we can see that more arid and/or cold areas tend to have higher proportions of free-living fixation. Barren land has the lowest BNF as it has no symbiotic BNF but has free-living fixation in the form of cryptogamic crusts. Conversely, we can see that temperate and tropical forested areas generally have a higher proportion of symbiotic BNF. This would explain why NPP could be a good proxy for BNF if symbiotic is assumed to be the major contribution to BNF. The low amount of BNF for Savanna is difficult to confirm and the subjectivity of allocation between Evergreen Broadleaf Forest, Savanna, and Open Shrubland increases the uncertainty.

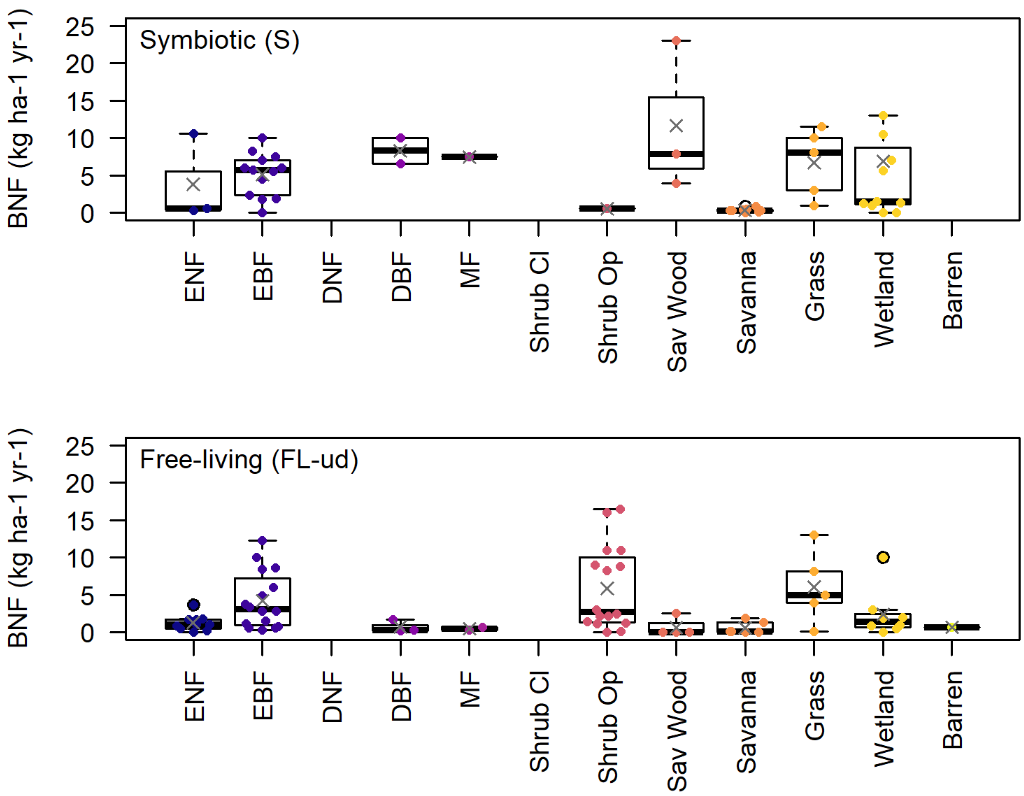


Figure 7. The symbiotic (S) and free-living (FL-ud) values categorised by biome type. For each boxplot, the midline is the median, the upper line third quartile, lower line the first

quartile, and the whiskers extend up to 1.5 times the interquartile range from the top of the box to the furthest datum within that distance. Datum beyond 1.5 times the interquartile range are represented as individual points. Overlaid on the boxplots are all the individual points as a ‘beeswarm’ scatter. The grey x on each set of data represents the arithmetic mean.

4 Discussion

Our central global estimate of 89 TgN yr^{-1} is only a little lower than most recent estimates (with the notable exception of Xu-Ri and Prentice, (2017)) (Figure 8). However, ours is one of only three estimates below 100 TgN yr^{-1} (the others being Delwiche, (1970) and Vitousek et al., (2013)). This represents a distinct lowering of the likely value of global BNF. The fact that Vitousek et al. (2013), using a completely independent top-down budget-method, proposes a low-end-range similar to ours gives more validity to our results. The upper end of our range encompasses most papers this century, but is vastly lower than Xu-Ri and Prentice, (2017) and Cleveland et al., (1999), which are well outside our range as well as being anomalous compared to all other global estimates. The large range emphasises the continuing uncertainty of global BNF values.

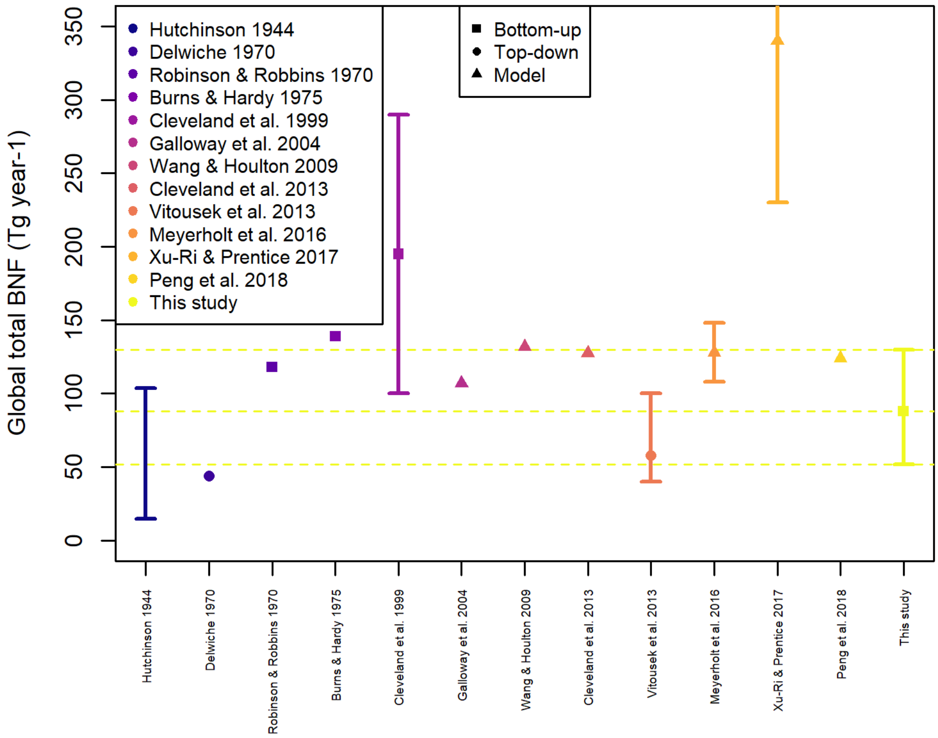


Figure 8. Global estimates of BNF, ordered by the publication date of the paper, plus the results from this study. Some model estimates are included for completeness, but most are a combination of measured data and modelling. Papers cited are: Burns & Hardy, (2012); Cleveland et al., (1999), (2013); Delwiche, (1970); Galloway et al., (2004); Hutchinson, (1944); Meyerholt et al., (2016); Peng et al., (2018); Robinson & Robbins, (1970); Vitousek et al., (2013); Wang Ying-Ping & Houlton Benjamin Z., (2009); Xu-Ri & Prentice, (2017). The symbols and categories relate to those identified in the introduction. The 'Model' category encompasses model only and a variety of model-data combinations. This figure shows all the observation-based values found in the literature and a representative selection of the modelled BNF values.

Few previous studies' estimates consider the relative contribution of free-living BNF, but one meta-analysis of BNF from cryptogamic crusts estimates it to be as much as 49 TgN yr^{-1} (Elbert et al., 2012). Our calculation of all free-living BNF (which encompasses cryptogamic crusts as well as other free-living BNF) is more modest at just 31 TgN yr^{-1} (Table 4), but still accounts for 36% of global BNF. Wang and Houlton, (2009) estimate 17 – 44% free-living in the tropical and extra-tropical regions respectively, broadly in line with what we found. Cleveland et al., (1999) does not explicitly state a ratio of free-living to symbiotic fixation, but their numbers suggest the free-living percentage is low. A later paper using similar data found free-living BNF accounted for only 18% of global BNF (Cleveland et al., 2013). Our study therefore suggests free-living fixation is a substantial contributor to BNF, possibly higher than previously thought. And while it remains dubious how helpful the symbiotic - free-living dichotomy is, there do appear to be important spatial and process differences between BNF types.

There is a slight indication from our statistical modelling that soil properties could be a determinant of BNF. However, without many more field measurements the number of values is not enough to do multivariate modelling. SOC, Soil Phosphorus, and Clay and Sand could all help predict BNF according to our analysis. Soil properties are known to be important to nitrogen limitation, as for instance young tropical soils are more nitrogen limited than old tropical soils (LeBauer & Treseder, 2008). If BNF were related to N limitation as is hypothesised in some models, the relative global homogeneity of N limitation (LeBauer & Treseder, 2008) would be consistent with our finding of BNF not having any strong global pattern.

Soil properties provide a possible predictor of BNF with some theoretical basis. Molybdenum and phosphorus availability are both known to affect BNF (Barron et al., 2009; Reed et al., 2007). Similarly, older soils tend to have higher carbon content and thus soil N also increases in accordance with a well-constrained global soil C:N ratio of 186:13 (Cleveland & Liptzin, 2007). Although organic N is not bioavailable to plants, mineralisation (organic to inorganic) of N is a significant contributor of N in modelled simulations of the N cycle (Zaehle, et al., 2014), even though as a proportion of total soil N it is low. N mineralisation makes up more than half of the N inputs into the global terrestrial N system according to an analysis done by Cleveland et al., (2013). There has also been research suggesting the role of mycorrhizal fungi in recycling N could be larger than previously thought (Terrer et al., 2016). Therefore, SOC could be a proxy for potentially available N. Because of the relative energetic costs, it stands to reason that as available soil N increases, N_2 fixation would decrease. It also is consistent with the theory of BNF being primarily an early succession feature of biomes, as SOC and other soil nutrients would be low at that point.

The question naturally arises why our results are at odds with the neat outcomes of Cleveland et al., (1999), who found a strong positive relationship with ET and NPP. There are two key differences that account for the discrepancy: the increase in available data, and methodology differences (particularly the separation of modelling and land cover type averaged upscaling in this study). A potentially useful context for the discrepancy of results would be how robust Cleveland et al., (1999)'s analysis is and thus how similar additional results would be expected to be. For further analysis of this, see Supporting Information, section 2.

The lack of a relationship between BNF and productivity at a macro scale shown by these results is in contrast to agricultural systems, where legume productivity is related to total BNF (Herridge et al., 2008). However, the difference between plant scale processes and ecosystem processes can be significant. Natural ecosystems would be expected to respond differently to the single plant scale or an agricultural system because the amount of fixers is variable and determined by natural selection and competition, rather than agricultural choice.

Since fixation has high energetic requirements, in most ecosystems non-fixers are more competitive and thus cover more area. In high productivity environments, such as tropical forest, non-fixers are therefore the main source of that high productivity. Even if symbiotic fixation at the plant level were higher in high than low productivity environments, they might still be out competed by non-fixers. In addition, it's not clear that the multitude of different types of free-living BNF organisms have the same relationship between fixation and productivity as seen in symbiotic agricultural species. Therefore, that natural ecosystem BNF does not scale with productivity simply shows the differences between natural and agricultural ecosystems.

One of the issues of low BNF estimations is difficulty in closing the global N budget. We know that new productivity requires nitrogen to maintain carbon-nitrogen ratios, and since inorganic N is soluble there are losses from the terrestrial biosphere. High estimates of BNF have sometimes been used as a convenient way to reconcile the apparent N shortage. This issue is muddied by the fact that many BNF estimations, especially from models, group together all non-deposition sources of new terrestrial N. Houlton et al., (2008) estimate N from weathering between 14 – 40 TgN yr^{-1} from denudation and 3 – 23 TgN yr^{-1} from chemical. Agricultural BNF has been calculated as 50 – 70 Tg $year^{-1}$ (Herridge et al., 2008) and analysis done by Vitousek et al., (2013) found that fertiliser from the Haber-Bosch process accounts for 120 Tg N yr^{-1}. The contribution of lightning to the nitrogen budget is thought to be small, around 7 Tg N yr-1 (Tie et al., 2002). Cumulatively, these could make a terrestrial N inputs large. However, the spatial distribution is very different between BNF and other terrestrial N sources and model developers need to be wary of assuming a single simple equation can capture the heterogeneity of either BNF or total terrestrial N inputs.

An alternative hypothesis for how low N input from BNF could be possible is that the N is not new but recycled. This could explain the incorrect paradigm of high BNF in the tropics by high nutrient recycling in the tropics. Terrer et al., (2016) suggests that mycorrhizal fungi could be responsible for much higher levels of N cycling than previously thought, and that ectomycorrhizal fungi and arbuscular mycorrhizal fungi could have different abilities to acquire N. The sample size of this study is small and has attracted considerable critique, however mycorrhizal fungi in principle could be a process that accounts for low levels of BNF where N limitation is also low.

The limitations of this analysis are mainly in the quantity of measurements available. However, quality and reliability are also key. Since N_2 is the most common gas in the atmosphere, small changes are difficult to measure accurately and precisely. As discussed in the methods, the measurement of N uptake via the acetylene-ethene assay reduction method (Hardy et al., 1968) is still prevalent despite margin for error in the conversion ratio (Nohrstedt, 1985; Saiz et al., 2019). Until better methodologies are cheaply and widely available, and enough new measurements are available to give reliable sample sizes and good global coverage, this is the best analysis possible.

5 Conclusions

Upscaling available symbiotic and free-living measured values obtained from natural ecosystems, by land cover, we estimate a median global value of 88 TgN yr^{-1} (52 – 130 TgN yr^{-1}) for BNF in natural terrestrial ecosystems. Our bottom-up estimate supports previous top-down methods that show a similarly low estimate of BNF. We found that at least a third of

BNF comes from free-living sources. In contradiction to previous work by Cleveland et al., (1999), we found no evidence for any relationship between BNF and either evapotranspiration or terrestrial productivity (NPP or GPP).

More field measurements are critical to progression of our understanding of BNF. The enormous heterogeneity of BNF at every level, especially in poorly represented areas such as Russia, Australia, Africa, and South East Asia make estimates uncertain. Multi-year field studies across several biomes are particularly rare at present. There is also a risk of null results (of no BNF being found) being left unpublished, even though absence of BNF is an important result. We urge the nitrogen community to continue to make BNF measurements, despite the seemingly large number already available, because without many more measurements with improved spatial and temporal distribution we cannot establish a more precise benchmark of BNF.

These datasets have a range of potential usages. The separate BNF type datasets and empirical models presented here open the possibility for modelling of free-living and symbiotic BNF in a more nuanced way than is presently done. Statistical modelling suggests soil characteristics show the most potential for an empirical relationship with BNF, which could theoretically be useful for models. The S and FL-ud BNF single categories are most important in terms of usefulness in projections and have the most measurements available. However, they are also poorest predicted. This presents a considerable challenge for modelling efforts. Therefore, the spatially identified maps of BNF provide the most opportunity by opening the possibility of comparing models to direct observational data.

Acknowledgements

Data used to generate the figures, plots, and tables in this paper can be acquired from the references in Supporting Information, section 1.

Authors acknowledge funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 641816 Coordinated Research in Earth Systems and Climate: Experiments kNowledge, Dissemination and Outreach (CRESCENDO).

References

Adams, D. G., & Duggan, P. S. (2008). Cyanobacteria–bryophyte symbioses. *Journal of Experimental Botany*, *59*(5), 1047–1058. https://doi.org/10.1093/jxb/ern005

Baker, T. G., Oliver, G. R., & Hodgkiss, P. D. (1986). Distribution and cycling of nutrients in Pinus radiata as affected by past lupin growth and fertiliser. *Forest Ecology and Management*, *17*(2), 169–187. https://doi.org/10.1016/0378-1127(86)90110-6

Barron, A. R., Wurzburger, N., Bellenger, J. P., Wright, S. J., Kraepiel, A. M. L., & Hedin, L. O. (2009). Molybdenum limitation of asymbiotic nitrogen fixation in tropical forest soils. *Nature Geoscience*, *2*(1), 42–45. https://doi.org/10.1038/ngeo366

Bauters, M., Mapenzi, N., Kearsley, E., Vanlauwe, B., & Boeckx, P. (2016). Facultative nitrogen fixation by legumes in the central Congo basin is downregulated during late successional stages. *Biotropica*, *48*(3), 281–284. https://doi.org/10.1111/btp.12312

Bloh, W. von, Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaehle, S. (2018). Implementing the nitrogen cycle into the dynamic global vegetation, hydrology, and crop growth model LPJmL (version 5.0). *Geoscientific Model Development*, *11*(7), 2789–2812. https://doi.org/10.5194/gmd-11-2789-2018

Blundon, D. J., & Dale, M. R. T. (1990). Dinitrogen Fixation (Acetylene Reduction) in Primary Succession Near Mount Robson, British Columbia, Canada. *Arctic and Alpine Research*, *22*(3), 255–263. https://doi.org/10.1080/00040851.1990.12002789

Bowman, W. D., Schardt, J. C., & Schmidt, S. K. (1996). Symbiotic N2-fixation in alpine tundra: ecosystem input and variation in fixation rates among communities. *Oecologia*, *108*(2), 345–350. https://doi.org/10.1007/BF00334660

Burns, R. C., & Hardy, R. W. F. (2012). *Nitrogen Fixation in Bacteria and Higher Plants*. Springer Science & Business Media.

Cech, P. G., Edwards, P. J., & Venterink, H. O. (2010). Why is Abundance of Herbaceous Legumes Low in African Savanna? A Test with Two Model Species. *Biotropica*, *42*(5), 580–589. https://doi.org/10.1111/j.1744-7429.2009.00622.x

Cleveland, C. C., & Liptzin, D. (2007). C:N:P stoichiometry in soil: is there a “Redfield ratio” for the microbial biomass? *Biogeochemistry*, *85*(3), 235–252. https://doi.org/10.1007/s10533-007-9132-0

Cleveland, C. C., Townsend, A. R., Schimel, D. S., Fisher, H., Howarth, R. W., Hedin, L. O., et al. (1999). Global patterns of terrestrial biological nitrogen (N2) fixation in natural ecosystems. *Global Biogeochemical Cycles*, *13*(2), 623–645. https://doi.org/10.1029/1999GB900014

Cleveland, C. C., Houlton, B. Z., Smith, W. K., Marklein, A. R., Reed, S. C., Parton, W., et al. (2013). Patterns of new versus recycled primary production in the terrestrial biosphere. *Proceedings of the National Academy of Sciences*, *110*(31), 12733–12737. https://doi.org/10.1073/pnas.1302768110

Davies-Barnard, T., Valdes, P. J., Singarayer, J. S., Wiltshire, A. J., & Jones, C. D. (2015). Quantifying the relative importance of land cover change from climate and land use in the representative concentration pathways. *Global Biogeochemical Cycles*, 2014GB004949. https://doi.org/10.1002/2014GB004949

Davis, M. (2014). Nitrogen leaching losses from forests in New Zealand. *New Zealand Journal of Forestry Science*, *44*, 2. https://doi.org/10.1186/1179-5395-44-2

Delwiche, C. C. (1970). The Nitrogen Cycle. *Scientific American*, *223*(3), 136–147.

Elbert, W., Weber, B., Burrows, S., Steinkamp, J., Büdel, B., Andreae, M. O., & Pöschl, U. (2012). Contribution of cryptogamic covers to the global cycles of carbon and nitrogen. *Nature Geoscience*, *5*(7), 459–462. https://doi.org/10.1038/ngeo1486

Fahey, T. J., Yavitt, J. B., Pearson, J. A., & Knight, D. H. (1985). The nitrogen cycle in lodgepole pine forests, southeastern Wyoming. *Biogeochemistry*, *1*(3), 257–275. https://doi.org/10.1007/BF02187202

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, 2001-2012, Collection 5.1 IGBP Land Cover,. Boston University, Boston, MA, USA. Retrieved from http://glcf.umd.edu/data/lc/

Galloway, J. N., Dentener, F. J., Capone, D. G., Boyer, E. W., Howarth, R. W., Seitzinger, S. P., et al. (2004). Nitrogen Cycles: Past, Present, and Future. *Biogeochemistry*, *70*(2), 153–226. https://doi.org/10.1007/s10533-004-0370-0

Goll, D. S., Winkler, A. J., Raddatz, T., Dong, N., Prentice, I. C., Ciais, P., & Brovkin, V. (2017). Carbon–nitrogen interactions in idealized simulations with JSBACH (version 3.10). *Geoscientific Model Development*, *10*(5), 2009–2030. https://doi.org/10.5194/gmd-10-2009-2017

Granhall, U. (1981). Biological Nitrogen Fixation in Relation to Environmental Factors and Functioning of Natural Ecosystems. *Ecological Bulletins*, (33), 131–144.

Grove, T. S., & Malajczuk, N. (1992). Nodule Production and Nitrogen Fixation (Acetylene Reduction) by an Understorey Legume (Bossiaea Laidlawiana) in Eucalyptus Forest. *Journal of Ecology*, *80*(2), 303–314. https://doi.org/10.2307/2261013

Hardy, R. W. F., Holsten, R. D., Jackson, E. K., & Burns, R. C. (1968). The Acetylene-Ethylene Assay for N2 Fixation: Laboratory and Field Evaluation. *Plant Physiology*, *43*(8), 1185–1207. https://doi.org/10.1104/pp.43.8.1185

Herridge, D. F., Peoples, M. B., & Boddey, R. M. (2008). Global inputs of biological nitrogen fixation in agricultural systems. *Plant and Soil*, *311*(1–2), 1–18. https://doi.org/10.1007/s11104-008-9668-3

Houlton, B. Z., Wang, Y.-P., Vitousek, P. M., & Field, C. B. (2008). A unifying framework for dinitrogen fixation in the terrestrial biosphere. *Nature*, *454*(7202), 327–330. https://doi.org/10.1038/nature07028

Hutchinson, G. E. (1944). Nitrogen in the Biogeochemistry of the Atmosphere. *American Scientist*, *32*(3), 178–195.

Johnson, H. B., & Mayeux, H. S. (1990). Prosopis glandulosa and the nitrogen balance of rangelands: extent and occurrence of nodulation. *Oecologia*, *84*(2), 176–185. https://doi.org/10.1007/BF00318269

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., et al. (2017). Compensatory water effects link yearly global land CO_2 sink changes to temperature. *Nature*, *541*(7638), 516–520. https://doi.org/10.1038/nature20780

Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., et al. (2013). The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. *Biogeosciences*, *10*(11), 7109–7131. https://doi.org/10.5194/bg-10-7109-2013

Kummerow, J., Alexander, J. V., Neel, J. W., & Fishbeck, K. (1978). Symbiotic Nitrogen Fixation in Ceanothus Roots. *American Journal of Botany*, *65*(1), 63–69. https://doi.org/10.1002/j.1537-2197.1978.tb10836.x

Le Quéré, C. L., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., et al. (2018). Global Carbon Budget 2018. *Earth System Science Data*, *10*(4), 2141–2194. https://doi.org/10.5194/essd-10-2141-2018

LeBauer, D. S., & Treseder, K. K. (2008). Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology*, *89*(2), 371–379. https://doi.org/10.1890/06-2057.1

Lenhart, K., Weber, B., Elbert, W., Steinkamp, J., Clough, T., Crutzen, P., et al. (2015). Nitrous oxide and methane emissions from cryptogamic covers. *Global Change Biology*, *21*(10), 3889–3900. https://doi.org/10.1111/gcb.12995

Ley, R. E., & D’Antonio, C. M. (1998). Exotic grass invasion alters potential rates of N fixation in Hawaiian woodlands. *Oecologia*, *113*(2), 179–187. https://doi.org/10.1007/s004420050366

Menge, D. N. L., & Chazdon, R. L. (2016). Higher survival drives the success of nitrogen-fixing trees through succession in Costa Rican rainforests. *New Phytologist*, *209*(3), 965–977. https://doi.org/10.1111/nph.13734

Menge, D. N. L., Levin, S. A., & Hedin, L. O. (2009). Facultative versus Obligate Nitrogen Fixation Strategies and Their Ecosystem Consequences. *The American Naturalist*, *174*(4), 465–477. https://doi.org/10.1086/605377

Meyerholt, J., Zaehle, S., & Smith, M. J. (2016). Variability of projected terrestrial biosphere responses to elevated levels of atmospheric CO_2 due to uncertainty in biological nitrogen fixation. *Biogeosciences*, *13*(5), 1491–1518. https://doi.org/10.5194/bg-13-1491-2016

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., et al. (2013). Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. *Hydrology and Earth System Sciences*, *17*(10), 3707–3720. https://doi.org/10.5194/hess-17-3707-2013

Nohrstedt, H.-Ö. (1985). Nonsymbiotic nitrogen fixation in the topsoil of some forest stands in central Sweden. *Canadian Journal of Forest Research*, *15*(4), 715–722. https://doi.org/10.1139/x85-116

Parkin, T. B., & Robinson, J. A. (1993). Statistical Evaluation of Median Estimators for Lognormally Distributed Variables. *Soil Science Society of America Journal*, *57*(2), 317–323. https://doi.org/10.2136/sssaj1993.03615995005700020005x

Peng, J., Dan, L., Wang, Y.-P., Tang, X., Yang, X., Yang, F., et al. (2018). How biological nitrogen fixation and climate change contribute to future terrestrial carbon sequestration from the global to biome scale. *Journal of Cleaner Production*. https://doi.org/10.1016/j.jclepro.2018.08.089

Permar, T. A., & Fisher, R. F. (1983). Nitrogen fixation and accretion by wax myrtle (Myrica cerifera) in slash pine (Pinus elliottii) plantations. *Forest Ecology and Management*, *5*(1), 39–46. https://doi.org/10.1016/0378-1127(83)90067-1

Porada, P., Weber, B., Elbert, W., Pöschl, U., & Kleidon, A. (2014). Estimating impacts of lichens and bryophytes on global biogeochemical cycles. *Global Biogeochemical Cycles*, *28*(2), 71–85. https://doi.org/10.1002/2013GB004705

Reed, S. C., Seastedt, T. R., Mann, C. M., Suding, K. N., Townsend, A. R., & Cherwin, K. L. (2007). Phosphorus fertilization stimulates nitrogen fixation and increases inorganic nitrogen concentrations in a restored prairie. *Applied Soil Ecology*, *36*(2), 238–242. https://doi.org/10.1016/j.apsoil.2007.02.002

Reed, S. C., Cleveland, C. C., & Townsend, A. R. (2011). Functional Ecology of Free-Living Nitrogen Fixation: A Contemporary Perspective. *Annual Review of Ecology, Evolution, and Systematics*, *42*(1), 489–512. https://doi.org/10.1146/annurev-ecolsys-102710-145034

Robinson, E., & Robbins, R. C. (1970). Gaseous Nitrogen Compound Pollutants from Urban and Natural Sources. *Journal of the Air Pollution Control Association*, *20*(5), 303–306. https://doi.org/10.1080/00022470.1970.10469405

Rundel, P. W., Nilsen, E. T., Sharifi, M. R., Virginia, R. A., Jarrell, W. M., Kohl, D. H., & Shearer, G. B. (1982). Seasonal dynamics of nitrogen cycling for a Prosopis woodland in the Sonoran Desert / Dinamica estacional del ciclo de nitrógeno de un bosque de Prosopis en el desierto Sonorense. *Plant and Soil*, *67*(1/3), 343–353.

Saiz, E., Sgouridis, F., Drijfhout, F. P., & Ullah, S. (2019). Biological nitrogen fixation in peatlands: Comparison between acetylene reduction assay and 15N2 assimilation methods. *Soil Biology and Biochemistry*, *131*, 157–165. https://doi.org/10.1016/j.soilbio.2019.01.011

Saxton, K. E., Rawls, W. J., Romberger, J. S., & Papendick, R. I. (1986). Estimating Generalized Soil-water Characteristics from Texture 1. *Soil Science Society of America Journal*, *50*(4), 1031–1036. https://doi.org/10.2136/sssaj1986.03615995005000040039x

Sheffer, E., Batterman, S. A., Levin, S. A., & Hedin, L. O. (2015). Biome-scale nitrogen fixation strategies selected by climatic constraints on nitrogen cycle. *Nature Plants*, *1*(12), 15182. https://doi.org/10.1038/nplants.2015.182

Smith, B., Wårlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., & Zaehle, S. (2014). Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. *Biogeosciences*, *11*(7), 2027–2054. https://doi.org/10.5194/bg-11-2027-2014

Terrer, C., Vicca, S., Hungate, B. A., Phillips, R. P., & Prentice, I. C. (2016). Mycorrhizal association as a primary control of the CO2 fertilization effect. *Science*, *353*(6294), 72–74. https://doi.org/10.1126/science.aaf4610

Tie, X., Zhang, R., Brasseur, G., & Lei, W. (2002). Global NOx Production by Lightning. *Journal of Atmospheric Chemistry*, *43*(1), 61–74. https://doi.org/10.1023/A:1016145719608

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., et al. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*, *13*(14), 4291–4313. https://doi.org/10.5194/bg-13-4291-2016

Vitousek, P. M., Menge, D. N. L., Reed, S. C., & Cleveland, C. C. (2013). Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1621), 20130119. https://doi.org/10.1098/rstb.2013.0119

Wang Ying-Ping, & Houlton Benjamin Z. (2009). Nitrogen constraints on terrestrial carbon uptake: Implications for the global carbon-climate feedback. *Geophysical Research Letters*, *36*(24). https://doi.org/10.1029/2009GL041009

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., et al. (2011). Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century. *Journal of Hydrometeorology*, *12*(5), 823–848. https://doi.org/10.1175/2011JHM1369.1

Wieder, W. R., Boehnert, J., & Bonan, G. B. (2014). Evaluating soil biogeochemistry parameterizations in Earth system models with observations. *Global Biogeochemical Cycles*, *28*(3), 211–222. https://doi.org/10.1002/2013GB004665

Xu-Ri, & Prentice, I. C. (2017). Modelling the demand for new nitrogen fixation by terrestrial ecosystems. *Biogeosciences*, *14*(7), 2003–2017. https://doi.org/10.5194/bg-14-2003-2017

Yang, X., Post, W. M., Thornton, P. E., & Jain, A. K. (2014). Global Gridded Soil Phosphorus Distribution Maps at 0.5-degree Resolution. *ORNL DAAC*. https://doi.org/10.3334/ORNLDAAC/1223

Zaehle, S., Medlyn, B. E., De Kauwe, M. G., Walker, A. P., Dietze, M. C., Hickler, T., et al. (2014). Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate Free-Air CO2 Enrichment studies. *New Phytologist*, *202*(3), 803–822. https://doi.org/10.1111/nph.12697

Zaehle, S., Jones, C. D., Houlton, B., Lamarque, J.-F., & Robertson, E. (2014). Nitrogen Availability Reduces CMIP5 Projections of Twenty-First-Century Land Carbon Uptake. *Journal of Climate*, *28*(6), 2494–2511. https://doi.org/10.1175/JCLI-D-13-00776.1